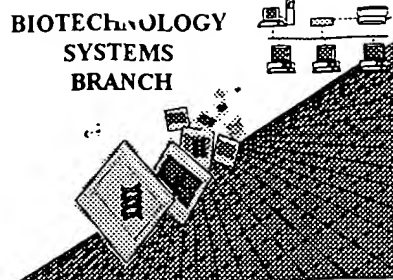


## RAW SEQUENCE LISTING ERROR REPORT



The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 09/671,635A

Source: 1645

Date Processed by STIC: 6/20/2001

RECEIVED

JUL 11 2001

TECH CENTER 1600/2900

**THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.**

**PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:**

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

**FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.**

**FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.**

**PATENTIN 2.1 e-mail help: [patin21help@uspto.gov](mailto:patin21help@uspto.gov) or phone 703-306-4119 (R. Wax)**

**PATENTIN 3.0 e-mail help: [patin3help@uspto.gov](mailto:patin3help@uspto.gov) or phone 703-306-4119 (R. Wax)**

**TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 3.0 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW:**

### **Checker Version 3.0**

The Checker Version 3.0 application is a state-of-the-art Windows based software program employing a logical and intuitive user-interface to check whether a sequence listing is in compliance with format and content rules. Checker Version 3.0 works for sequence listings generated for the original version of 37 CFR §§1.821 – 1.825 effective October 1, 1990 (old rules) and the revised version (new rules) effective July 1, 1998 as well as World Intellectual Property Organization (WIPO) Standard ST.25.

Checker Version 3.0 replaces the previous DOS-based version of Checker, and is Y2K-compliant. Checker allows public users to check sequence listings in Computer Readable form (CRF) before submitting them to the United States Patent and Trademark Office (USPTO). Use of Checker prior to filing the sequence listing is expected to result in fewer errored sequence listings, thus saving time and money.

**Checker Version 3.0 can be down loaded from the USPTO website at the following address:**

**<http://www.uspto.gov/web/offices/pac/checker>**

RECEIVED

JUL 11 2001

## Raw Sequence Listing Error Summary

TECH CENTER 1600/2900

ERROR DETECTED      SUGGESTED CORRECTIONSERIAL NUMBER: 09/674,635A

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1      Wrapped Nucleics      The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."
- 1      Wrapped Aminos
- 2      Invalid Line Length      The rules require that a line not exceed 72 characters in length. This includes white spaces.
- 3      Misaligned Amino      The numbering under each 5<sup>th</sup> amino acid is misaligned. Do not use tab codes between numbers; use space characters, instead.
- 3      Numbering
- 4      Non-ASCII      The submitted file was not saved in ASCII(DOS) text, as required by the Sequence Rules. Please ensure your subsequent submission is saved in ASCII text.
- 5      Variable Length      Sequence(s)      contain n's or Xaa's representing more than one residue. Per Sequence Rules, each n or Xaa can only represent a single residue. Please present the maximum number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.
- 6      PatentIn 2.0      A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s)     . Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.
- 6      "bug"
- 7      Skipped Sequences      Sequence(s)      missing. If intentional, please insert the following lines for each skipped sequence:  
(2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
(i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading)  
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
This sequence is intentionally skipped  
  
Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to include the skipped sequences.
- 7      (OLD RULES)
- 8      Skipped Sequences      Sequence(s)      missing. If intentional, please insert the following lines for each skipped sequence.  
<210> sequence id number  
<400> sequence id number  
000
- 8      (NEW RULES)
- 9      Use of n's or Xaa's      Use of n's and/or Xaa's have been detected in the Sequence Listing.  
(NEW RULES)      Per 1.823 of Sequence Rules, use of <220>-<223> is MANDATORY if n's or Xaa's are present. In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.
- 10      Invalid <213>      Per 1.823 of Sequence Rules, the only valid <213> responses are: Unknown, Artificial Sequence, or  
Response      scientific name (Genus/species). <220>-<223> section is required when <213> response is Unknown or is Artificial Sequence
- 11      Use of <220>      Sequence(s)      missing the <220> "Feature" and associated numeric identifiers and responses. Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section.  
(See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)
- 12      PatentIn 2.0      Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.
- 12      "bug"

RECEIVED

JUL 11 2001

1645

TECH CENTER 1600/2900

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/671,635A

DATE: 06/20/2001

TIME: 11:08:56

Input Set : N:\COPIES\ES.txt

Output Set: N:\CRF3\06202001\I671635A.raw

Does Not Comply  
Corrected Diskette Needed

3 <110> APPLICANT: ALEXANDROV, Nickolai et al.  
5 <120> TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING  
POLYPEPTIDES ENCODED

6 THEREBY

8 &lt;130&gt; FILE REFERENCE: 2750-1026P

10 &lt;140&gt; CURRENT APPLICATION NUMBER: US 09/671,635A

11 &lt;141&gt; CURRENT FILING DATE: 2000-09-28

13 &lt;160&gt; NUMBER OF SEQ ID NOS: 802

15 &lt;170&gt; SOFTWARE: PatentIn version 3.0

17 &lt;210&gt; SEQ ID NO: 1

18 &lt;211&gt; LENGTH: 4

19 &lt;212&gt; TYPE: PRT

20 &lt;213&gt; ORGANISM: Consensus Sequence

22 &lt;400&gt; SEQUENCE: 1

24 Ala Gly Cys Asn

25 1

27 &lt;210&gt; SEQ ID NO: 2

28 &lt;211&gt; LENGTH: 4

29 &lt;212&gt; TYPE: PRT

30 &lt;213&gt; ORGANISM: Consensus Sequence

32 &lt;400&gt; SEQUENCE: 2

34 Ala Gly Ile Met

35 1

37 &lt;210&gt; SEQ ID NO: 3

38 &lt;211&gt; LENGTH: 4

39 &lt;212&gt; TYPE: PRT

40 &lt;213&gt; ORGANISM: Consensus Sequence

42 &lt;400&gt; SEQUENCE: 3

44 Ala Gly Leu Ile

45 1

47 &lt;210&gt; SEQ ID NO: 4

48 &lt;211&gt; LENGTH: 4

49 &lt;212&gt; TYPE: PRT

50 &lt;213&gt; ORGANISM: Consensus Sequence

52 &lt;400&gt; SEQUENCE: 4

54 Ala Gly Leu Met

55 1

57 &lt;210&gt; SEQ ID NO: 5

58 &lt;211&gt; LENGTH: 5

59 &lt;212&gt; TYPE: PRT

60 &lt;213&gt; ORGANISM: Consensus Sequence

62 &lt;400&gt; SEQUENCE: 5

64 Ala Gly Ser Cys Ile

65 1

67 &lt;210&gt; SEQ ID NO: 6

68 &lt;211&gt; LENGTH: 5

69 &lt;212&gt; TYPE: PRT

70 &lt;213&gt; ORGANISM: Consensus Sequence

see item 10 on Error Summary sheet  
(global error)

## RAW SEQUENCE LISTING

DATE: 06/20/2001

PATENT APPLICATION: US/09/671,635A

TIME: 11:08:56

Input Set : N:\COPIES\ES.txt

Output Set: N:\CRF3\06202001\I671635A.raw

72 <400> SEQUENCE: 6  
74 Ala Gly Ser Asp Met  
75 1 5  
77 <210> SEQ ID NO: 7  
78 <211> LENGTH: 4  
79 <212> TYPE: PRT  
80 <213> ORGANISM: Consensus Sequence  
82 <400> SEQUENCE: 7  
84 Ala Ile Val Pro  
85 1  
87 <210> SEQ ID NO: 8  
88 <211> LENGTH: 4  
89 <212> TYPE: PRT  
90 <213> ORGANISM: Consensus Sequence  
92 <400> SEQUENCE: 8  
94 Ala Leu Ile Val  
95 1  
97 <210> SEQ ID NO: 9  
98 <211> LENGTH: 4  
99 <212> TYPE: PRT  
100 <213> ORGANISM: Consensus Sequence  
102 <400> SEQUENCE: 9  
104 Ala Pro Asn Thr  
105 1  
107 <210> SEQ ID NO: 10  
108 <211> LENGTH: 5  
109 <212> TYPE: PRT  
110 <213> ORGANISM: Consensus Sequence  
112 <400> SEQUENCE: 10  
114 Ala Ser Leu Val Arg  
115 1 5  
117 <210> SEQ ID NO: 11  
118 <211> LENGTH: 5  
119 <212> TYPE: PRT  
120 <213> ORGANISM: Consensus Sequence  
122 <400> SEQUENCE: 11  
124 Ala Ser Thr Asp Val  
125 1 5  
127 <210> SEQ ID NO: 12  
128 <211> LENGTH: 5  
129 <212> TYPE: PRT  
130 <213> ORGANISM: Consensus Sequence  
132 <400> SEQUENCE: 12  
134 Ala Ser Thr Pro Val  
135 1 5  
137 <210> SEQ ID NO: 13  
138 <211> LENGTH: 5  
139 <212> TYPE: PRT  
140 <213> ORGANISM: Consensus Sequence

## RAW SEQUENCE LISTING

DATE: 06/20/2001

PATENT APPLICATION: US/09/671,635A

TIME: 11:08:56

Input Set : N:\COPIES\ES.txt

Output Set: N:\CRF3\06202001\I671635A.raw

142 <400> SEQUENCE: 13  
144 Ala Val Asn His Lys  
145 1 5  
147 <210> SEQ ID NO: 14  
148 <211> LENGTH: 5  
149 <212> TYPE: PRT  
150 <213> ORGANISM: Consensus Sequence  
152 <400> SEQUENCE: 14  
154 Cys Ser Ala Gly Asn  
155 1 5  
157 <210> SEQ ID NO: 15  
158 <211> LENGTH: 4  
159 <212> TYPE: PRT  
160 <213> ORGANISM: Consensus Sequence  
162 <400> SEQUENCE: 15  
164 Cys Ser Ala Met  
165 1  
167 <210> SEQ ID NO: 16  
168 <211> LENGTH: 4  
169 <212> TYPE: PRT  
170 <213> ORGANISM: Consensus Sequence  
172 <400> SEQUENCE: 16  
174 Cys Ser Ala Val  
175 1  
177 <210> SEQ ID NO: 17  
178 <211> LENGTH: 4  
179 <212> TYPE: PRT  
180 <213> ORGANISM: Consensus Sequence  
182 <400> SEQUENCE: 17  
184 Cys Ser Thr Ala  
185 1  
187 <210> SEQ ID NO: 18  
188 <211> LENGTH: 6  
189 <212> TYPE: PRT  
190 <213> ORGANISM: Consensus Sequence  
192 <400> SEQUENCE: 18  
194 Cys Ser Thr Ala Glu Asn  
195 1 5  
197 <210> SEQ ID NO: 19  
198 <211> LENGTH: 5  
199 <212> TYPE: PRT  
200 <213> ORGANISM: Consensus Sequence  
202 <400> SEQUENCE: 19  
204 Asp Ala Gly His Glu  
205 1 5  
207 <210> SEQ ID NO: 20  
208 <211> LENGTH: 4  
209 <212> TYPE: PRT  
210 <213> ORGANISM: Consensus Sequence

## RAW SEQUENCE LISTING

DATE: 06/20/2001

PATENT APPLICATION: US/09/671,635A

TIME: 11:08:56

Input Set : N:\COPIES\ES.txt

Output Set: N:\CRF3\06202001\I671635A.raw

212 <400> SEQUENCE: 20  
214 Asp Glu Ala Gly  
215 1  
217 <210> SEQ ID NO: 21  
218 <211> LENGTH: 4  
219 <212> TYPE: PRT  
220 <213> ORGANISM: Consensus Sequence  
222 <400> SEQUENCE: 21  
224 Asp Glu Ala Pro  
225 1  
227 <210> SEQ ID NO: 22  
228 <211> LENGTH: 5  
229 <212> TYPE: PRT  
230 <213> ORGANISM: Consensus Sequence  
232 <400> SEQUENCE: 22  
234 Asp Glu Phe Tyr Trp  
235 1 5  
237 <210> SEQ ID NO: 23  
238 <211> LENGTH: 8  
239 <212> TYPE: PRT  
240 <213> ORGANISM: Consensus Sequence  
242 <400> SEQUENCE: 23  
244 Asp Glu Gly Ser Thr His Lys Arg  
245 1 5  
247 <210> SEQ ID NO: 24  
248 <211> LENGTH: 8  
249 <212> TYPE: PRT  
250 <213> ORGANISM: Consensus Sequence  
252 <400> SEQUENCE: 24  
254 Asp Glu Lys Arg His Ser Thr Ala  
255 1 5  
257 <210> SEQ ID NO: 25  
258 <211> LENGTH: 4  
259 <212> TYPE: PRT  
260 <213> ORGANISM: Consensus Sequence  
262 <400> SEQUENCE: 25  
264 Asp Glu Asn Gly  
265 1  
267 <210> SEQ ID NO: 26  
268 <211> LENGTH: 6  
269 <212> TYPE: PRT  
270 <213> ORGANISM: Consensus Sequence  
272 <400> SEQUENCE: 26  
-----274 Asp Glu Asn Lys Ala Cys-----  
275 1 5  
277 <210> SEQ ID NO: 27  
278 <211> LENGTH: 6  
279 <212> TYPE: PRT  
280 <213> ORGANISM: Consensus Sequence

## RAW SEQUENCE LISTING

DATE: 06/20/2001


PATENT APPLICATION: US/09/671,635A

TIME: 11:08:56

Input Set : N:\COPIES\ES.txt

Output Set: N:\CRF3\06202001\I671635A.raw

282 <400> SEQUENCE: 27  
284 Asp Glu Asn Lys His Ser  
285 1 5  
287 <210> SEQ ID NO: 28  
288 <211> LENGTH: 5  
289 <212> TYPE: PRT  
290 <213> ORGANISM: Consensus Sequence  
292 <400> SEQUENCE: 28  
294 Asp Glu Asn Lys Ser  
295 1 5  
297 <210> SEQ ID NO: 29  
298 <211> LENGTH: 4  
299 <212> TYPE: PRT  
300 <213> ORGANISM: Consensus Sequence  
302 <400> SEQUENCE: 29  
304 Asp Glu Asn Leu  
305 1  
307 <210> SEQ ID NO: 30  
308 <211> LENGTH: 6  
309 <212> TYPE: PRT  
310 <213> ORGANISM: Consensus Sequence  
312 <400> SEQUENCE: 30  
314 Asp Glu Asn Pro His Lys  
315 1 5  
317 <210> SEQ ID NO: 31  
318 <211> LENGTH: 4  
319 <212> TYPE: PRT  
320 <213> ORGANISM: Consensus Sequence  
322 <400> SEQUENCE: 31  
324 Asp Glu Asn Gln  
325 1  
327 <210> SEQ ID NO: 32  
328 <211> LENGTH: 7  
329 <212> TYPE: PRT  
330 <213> ORGANISM: Consensus Sequence  
332 <400> SEQUENCE: 32  
334 Asp Glu Asn Gln Ala Arg Lys  
335 1 5  
337 <210> SEQ ID NO: 33  
338 <211> LENGTH: 6  
339 <212> TYPE: PRT  
340 <213> ORGANISM: Consensus Sequence  
342 <400> SEQUENCE: 33  
344 Asp-Glu-Asn-Gln-Gly-Ala  
345 1 5  
347 <210> SEQ ID NO: 34  
348 <211> LENGTH: 8  
349 <212> TYPE: PRT  
350 <213> ORGANISM: Consensus Sequence



The types of errors shown exist throughout the Sequence Listing. Please check subsequent sequences for similar errors.

VERIFICATION SUMMARY

DATE: 06/20/2001

PATENT APPLICATION: US/09/671,635A

TIME: 11:08:57

Input Set : N:\COPIES\ES.txt

Output Set: N:\CRF3\06202001\I671635A.raw